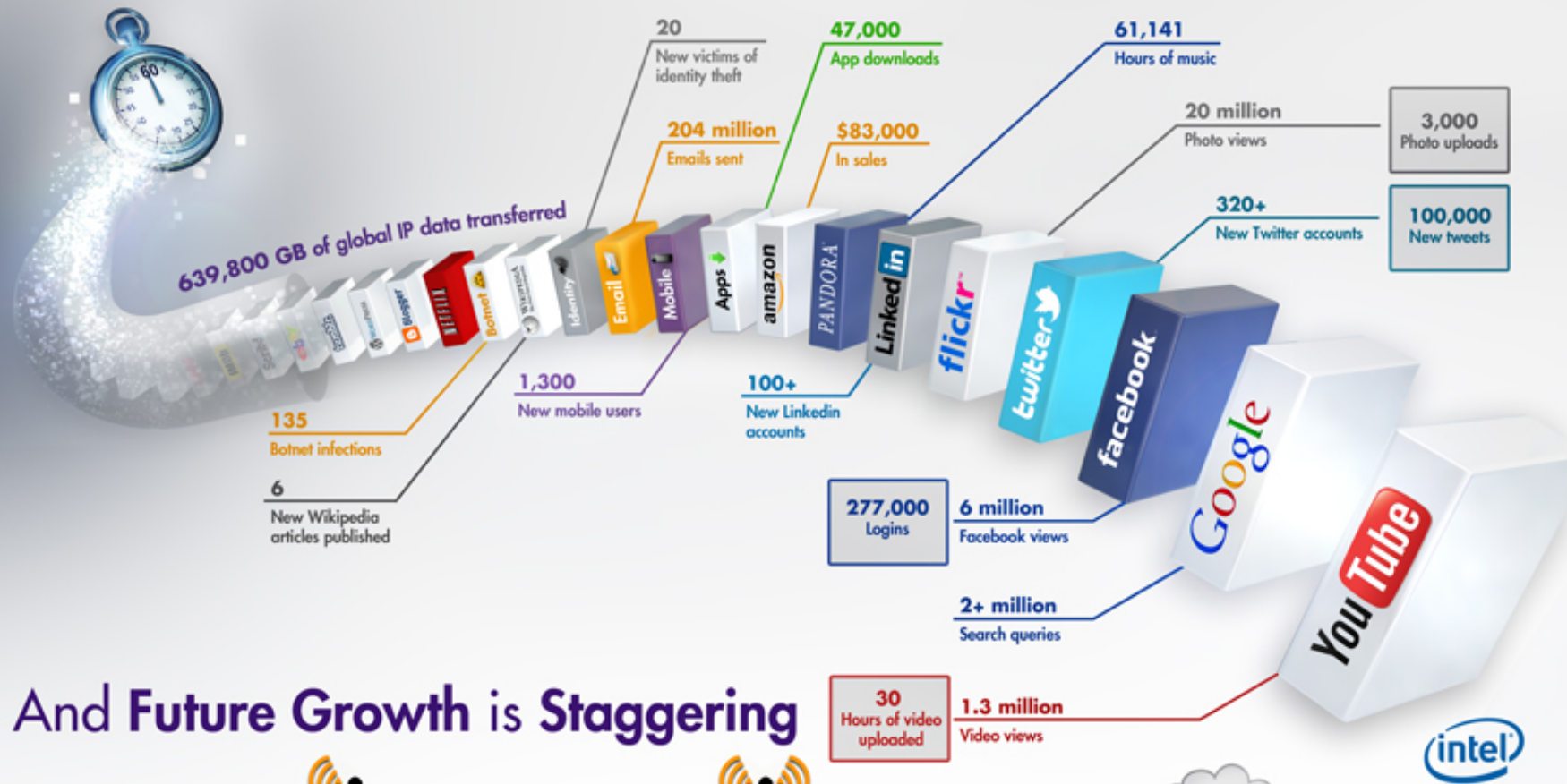


# What Happens in an Internet Minute?

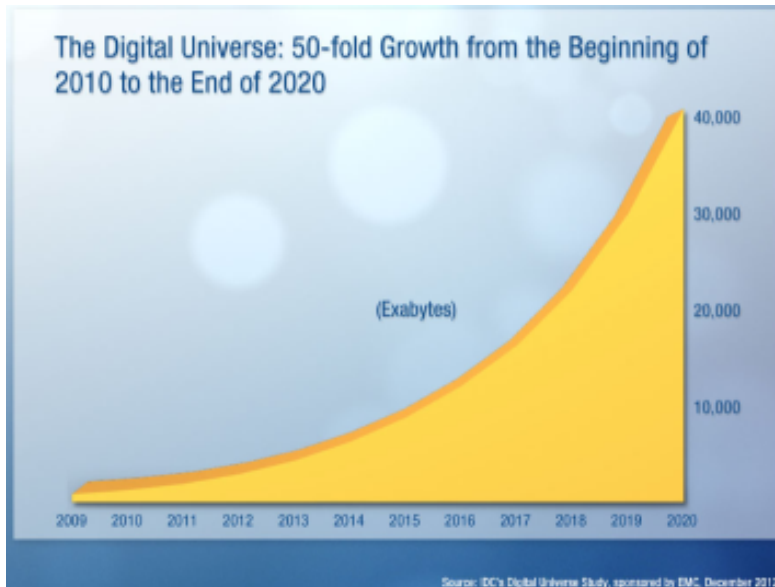


## And Future Growth is Staggering



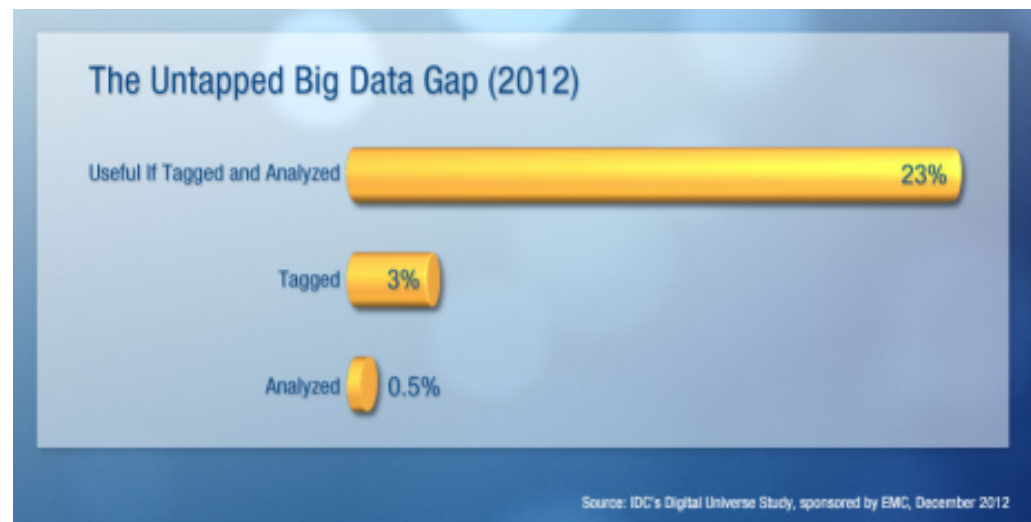
Source: [Intel](http://www.intel.com)

# Big Data Growing



IDC predicts: Through 2020, the digital universe will double every 2 years and grow from 130 exabytes to 40,000 exabytes or 5,200 GB / person in 2020.

**The Untapped Data Gap:**  
Most of the useful data  
will not be tagged or  
analyzed – partly due to  
skill shortage



# What's Seen the Most Recent Growth

Types of Data	Types of Activities/Areas
<ul style="list-style-type: none"><li>• Location / Geo / Mobile Data</li><li>• Music / Audio</li><li>• Social Media / Social Networks</li><li>• Time Series</li><li>• Images / Video</li><li>• User Profile data</li><li>• Text feeds / Micro-blog data</li></ul>	<ul style="list-style-type: none"><li>• Search / Web content mining</li><li>• Text mining / opinion analysis</li><li>• Personalization / recommendation</li><li>• Social network / media analysis</li><li>• Topic modeling / micro-blog analysis</li><li>• Health informatics</li></ul>

- Much of this growth is driven by end user mobile or Web-based applications
  - need for **more personalized and intelligent applications integrating data across many channels**

# With Big Data Challenges Come Big Data Opportunities

- Google – first big success of big data
- Social networks (Facebook, Twitter, LinkedIn, ...) success depends on network size, i.e. big data
- Integration across many channels and data types can lead to more discoveries; better solutions
  - Domain knowledge, collaborative data, behavior data, location/geo data, content data, social networks, sensor/observation data
- Better analytics in vertical domains
  - Health Analytics
  - Education
  - Hospitality/Travel
  - Scientific Computing

# Data as “New Oil”?



- Big Data: more like crude oil
  - messy, lots of contaminants, needs refinement
  - This is where data mining and analytics come in

# What's needed to succeed in the new world of “big data”?

- **Leveraging big data**

- Biggest challenge is in data distillation and integration across many channels

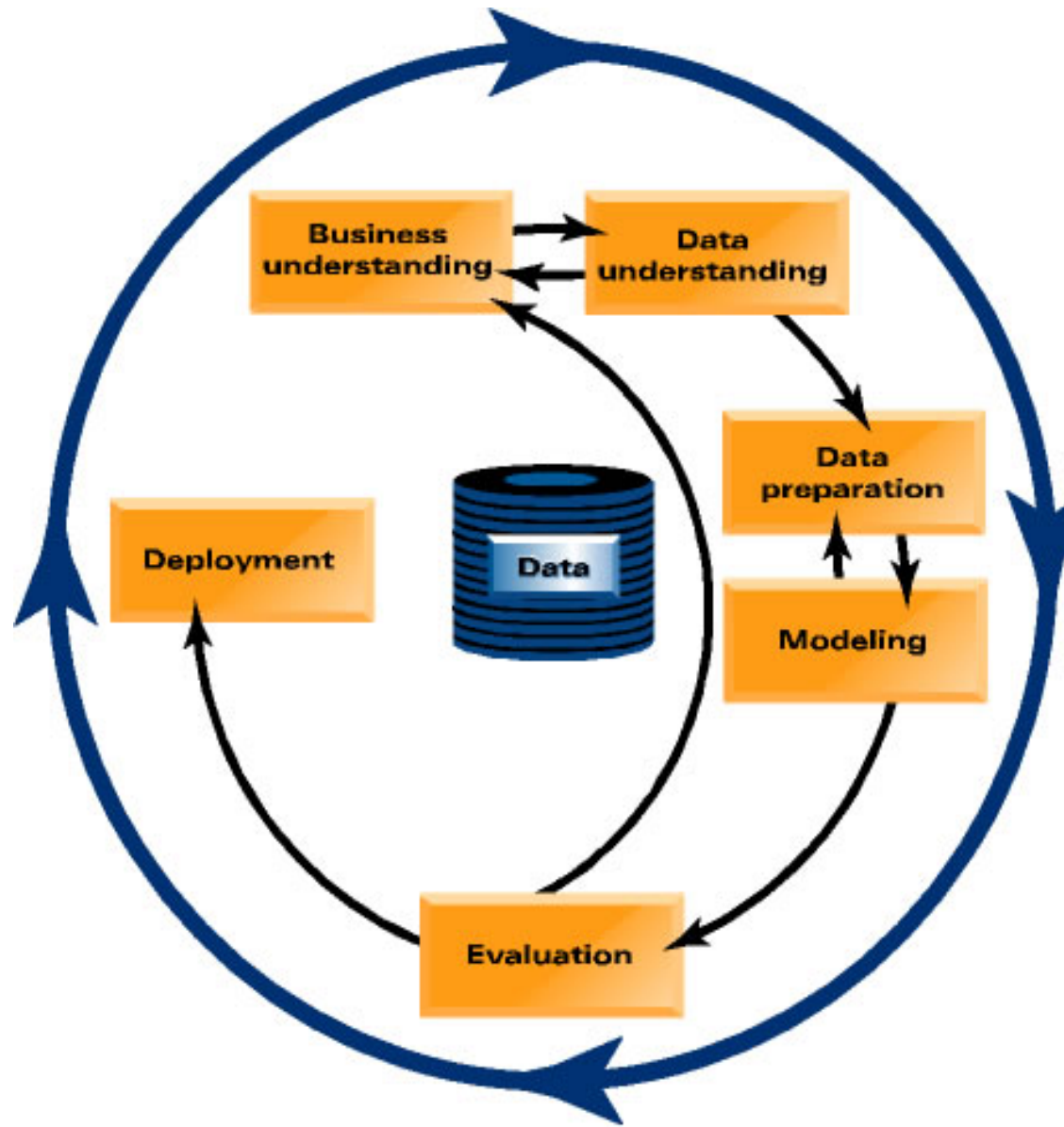
- **Effective use of analytics**

- Not just analytics, but predictive analytics
- No longer just a luxury but an integral part of systems

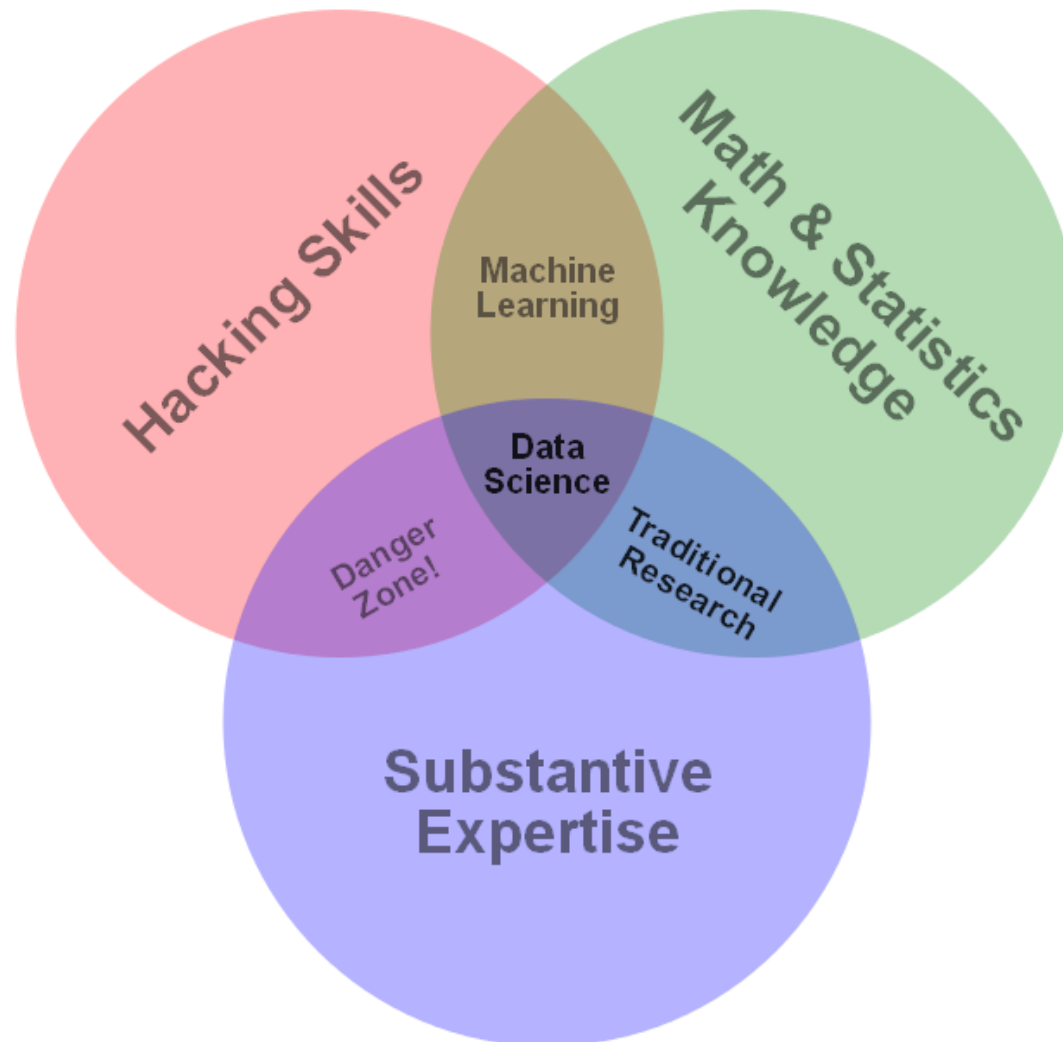
- **Real-time deployment of predictive models**

- Needed for effective delivery of relevant, targeted, personalized content





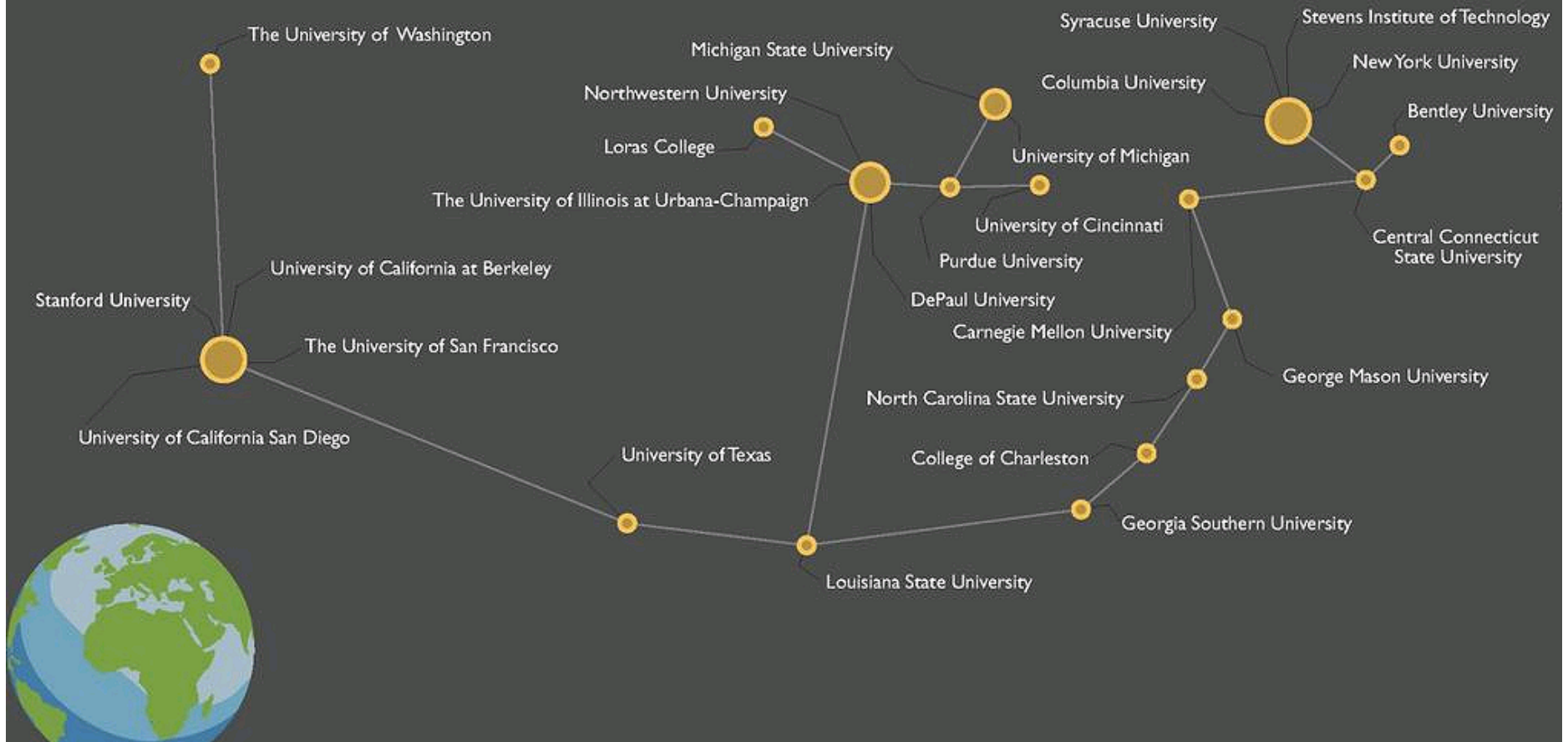
# Drew Conway's Data Science Venn Diagram





# A Constellation Is Born

*Data Science classes forming across the country*



# DePaul Links

- MS in Predictive Analytics:



<http://www.cdm.depaul.edu/academics/Pages/MS-in-Predictive-Analytics.aspx>

- Center for Data Mining and Predictive Analytics



<http://dampa.cdm.depaul.edu/index.aspx>